I.J. Image, Graphics and Signal Processing, **2025, 6, 90-108**
Published Online on December 8, 2025 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2025.06.06

Dual Attention Fusion-Net with Edge Attention Guidance Network based Segmentation for an Automatic Size Detection of Onions

M. Mythili*
Information Science & Engineering, Vemana Institute of Technology, Koramangala, Bengaluru- 560034, Karnataka, India
Email: mythi.m84@gmail.com
ORCID iD: https://orcid.org/0000-0002-0071-6178
*Corresponding Author

P. Vasanthi Kumari
Department of Computer Applications, Dayananda Sagar University (Innovation Campus), Singasandra, Bengaluru-560068, Karnataka, India
Email: vasanthi-bca@dsu.edu.in
ORCID iD: https://orcid.org/0000-0002-6063-8415

Received: 04 November, 2024; Revised: 20 May, 2025; Accepted: 15 September, 2025; Published: 08 December, 2025

Abstract: Onion size is a crucial physiological characteristic that can be explained by a number of factors, including diameter, weight, volume, and length. Determining the size of onions is frequently necessary for sorting them for a variety of reasons, including processing machine specifications, legal requirements for sorting standards, and consumer preferences. In the process of phenotyping onions, size is another crucial quantitative feature to consider. Traditionally, algorithms based on morphology, colour, thresholding, and geometric approaches have been used to estimate the shape and size of onions. However, research that relies on these geometric or colour-based functions is limited to approximations and frequently produces erroneous results when conducted at precisely controlled heights. Healthy onions are collected and utilized as an input dataset for this paper. The gathered images are pre-processed to reduce noise and improve contrast by applying the circular adaptive median filter and homomorphic filtering with Elk-herd optimization. Next, utilizing the dilated and deformable feature pyramid network, object detection is performed on the pre-processed images. To segment the onion from the image for removing the unwanted portions, an edge-based segmentation algorithm is used, such as an edge-attention guidance network. The dual attention fusion-net, which ranks data into labelled groups and measures onion size. Accuracy, confusion metrics, FDR, hit rate, and other performance metrics are assessed for both the current and proposed models in the proposed model. Consequently, the suggested onion size detection approach outperforms the current algorithm. This method produced 97.6% accuracy, 2.9% FDR, 96% Hit Rate, 98.5% Selectivity, and 97.3% NPV. Thus, this proposed approach is the best choice for detecting the size of the onion.

Index Terms: Onion Size Detection, Elk-Herd Optimization, Dual Attention Fusion-Net, Dilated & Deformable Feature Pyramid Network, Circular Adaptive Median Filter, Edge Attention Guidance Network

1. Introduction

Significant technological improvements have been applied throughout most of human history to increase agricultural production with limited resources [1]. Utilizing automation technology and IntelliSense, agricultural automation dramatically increases the productivity of agricultural production [2]. Recently, agricultural production operations have made substantial use of deep learning and computer vision (CV) [3]. With the use of CV, farmers can acquire a multitude of metrics and data regarding the crops through images. When it comes to real-time performance, unified standards, high efficiency, and good measurement, automation measurement is superior than manual measurement [4]. The assessment of internal quality (IQ) is now known to play a significant impact in the post-harvest stage of fruits and vegetables; fresh products external appearance will also continue to play a significant function; and

fruit size is a key component of external appearance. The ability to classify fresh market fruits into size categories makes fruit size determination crucial for a number of reasons. This makes it easy to match consumer preferences, determine the market and price differences between large and small produce, and possibly most importantly allow pattern packing [5]. Size determination is necessary for contemporary or dry online fruit and vegetable density sorting, which requires the volume and weight of two sized-related characteristics. Determining the products surface area based on IQ sensors can function properly [6].

Majority of current techniques for measuring the sizes of fruits and vegetables rely on basic image processing techniques like bounding boxes and edge detection. While these techniques are adequate for fruits with normal shapes, they are difficult to use for complicated ones [7]. Furthermore, these techniques typically require the camera distance from the fruit as presumptive data. They can only be classified as fruits once they have been picked. Certain works provide the depth value in order to finish measuring the vegetables size on the tree [8]. These techniques, however, are limited to certain kinds of fruits and vegetables. Initially, keypoint detection was suggested for estimating human stance [9]. Its positive impact on agricultural automation has been demonstrated by some recent research. An enhanced YOLO and RGB-D camera were used to create an earlier tomato peduncle identification technique for autonomous harvesting. Keypoint detection is usually used for autonomous fruit harvesting in order to identify and maintain. However, post-processing is still necessary to ascertain the fruit and vegetable sizes [10]. So, in this proposed approach a dual attention fusion net model is developed to predict the onion size as three different categories.

The suggested work's primary contribution is listed below:

- A novel deep learning approach is developed for effective segmentation and accurate prediction of an onion size.
- Circular adaptive median filter and homomorphic filter are used to remove the noise and improve the quality of raw healthy onion images.
- Reflectance component's ideal value is selected in the homomorphic filter using the Elk-Herd Optimization (EHO) optimization to improve the details in the onion image.
- For detecting the onion from the collected images, the Dilated and Deformable Feature Pyramid Network (DDFPN) is used.
- Edge-Attention Guidance Network (ET-Net) is utilized to segment the object-detected image, which recognizes and segments the image.
- Dual Attention Fusion-Network (DAF-Net) is employed to classify the accurate size of an onion from the segmented images with the area of an onion size.

The remaining sections of this work are as follows: Section 2 evaluates the reviews related to fruits and vegetables size detection; Section 3 describes the proposed methodology and its architecture; Section 4 presents the result of the proposed technique; and the Conclusion section is given in section 5.

2. Literature Review

This section examines most of the in-depth articles on the topic of vegetable and fruit detection using different algorithms; a few of them are discussed below along with their drawbacks.

Apolo-Apolo et al [11] established an automated image processing methods that uses deep learning techniques to detect the yield and size of citrus fruits. The images were acquired from an unmanned aerial vehicle (UAV). A Faster R-CNN Deep Learning model was trained to identify oranges in the acquired images. The encouraging outcomes show that size discrimination before harvest may be accomplished with this size estimation method. Nevertheless, as it is impossible to calculate the sizes of every fruit found, this method was generic in character.

Ponce et al [12] introduced a computer vision and feature modelling approach to enable accurate automatic olive-fruit grading. By statistically correlating the data obtained from image analysis with the respective reference measurements, models evaluated these features. The models were evaluated using 2700 external samples, yielding relative errors of the main and minor axis lengths for all types that were below 0.80% and 1.05%, respectively. Nevertheless, these techniques are limited to a certain kind of fruit.

Gene-mola et al [13] suggested a brand-new, four-step process for automatically estimating apple size in the field: Fruit detection, Multi-view stereo (MVS) and structure-from-motion (SfM) point cloud generation, fruit size estimate, and fruit visibility estimation were the first four tasks. Optimum results were obtained with the least squares, M−estimator sample consensus and template matching which obtained the coefficients of determination of 0.88, 0.91, and 0.88, respectively, and the mean absolute errors of 4.5 mm, 3.7 mm, and 4.2 mm. Large-scale agricultural fields were not able to directly use this technology due to its high processing time.

Jeong et al [14] suggested an AI technology which includes computer vision algorithms in conjunction with LiDAR sensor data to precisely determine the dimensions and weight of strawberries. By utilizing LiDAR sensor capabilities and combining deep learning models, to reduce the need for human interaction. The strawberries height and width had relative faults of 5.42% and 3.71% with the width showing a lower error rate. The standard deviations of width and

height was 0.19% and 0.24%, respectively. Nonetheless, the average relative error has a propensity to inflate by specific sizes with comparatively larger relative errors.

Ferrer-Ferrer et al [15] suggested an innovative Deep Neural Network-based method for measuring and detecting apples in the field. The proposed framework's end-to-end multitask Deep Neural Network architecture, which was trained using RGB-D data. The diameter of apples ranging from 27 mm to 95 mm showed that the mean absolute error of 5.64 mm, and the F1-score for apple detection was 0.88. The measurement was restricted to apples with visibility above 65%, the prediction performed better. However, the performance of these systems is affected by background complexity, motion blur and low light.

Sapkota et al [16] employed 3D point cloud data and cutting-edge YOLOv8 object identification and instance segmentation with geometric shape fitting algorithms to compute the exact size of immature green apples (or fruitlets). Estimated the size of apple fruitlets with root mean square error of 2.35 mm, mean absolute error of 1.66 mm, mean absolute percentage error of 6.15 mm, and an R-squared value of 0.9 using the ellipsoid fitting using images from the Azure Kinect. However, this study was limited to the glasshouse environment.

As per the previously cited research, several fruits and vegetable aspect-based recommendation algorithms now in use have several disadvantages. It is impossible to calculate the sizes of every fruit [11], it is limited to one kind of fruit [12], high processing time [13], larger relative errors [14], motion blur and low light [15], limited to the glasshouse environment [16]. To overcome these issues effective deep learning approaches are developed in this proposed approach to segment and predict the onion and its size.

3. Proposed Methodology

Developing methods for assessing onion size is necessary for agricultural automation technology. Onions are one of the most widely grown, highly traded, and nutritiously valuable vegetables. However, manual grading by visual inspection takes a lot of time and effort, and it could not produce consistent results. In addition, the multi-step procedure of hand grading may cause physical harm to onions and financial losses during export. Thus, computer vision offers an automated and quick way to complete these tasks. So in this proposed approach Dual Attention Fusion-Network (DAF-Net) is designed for predicting the size of an onion.

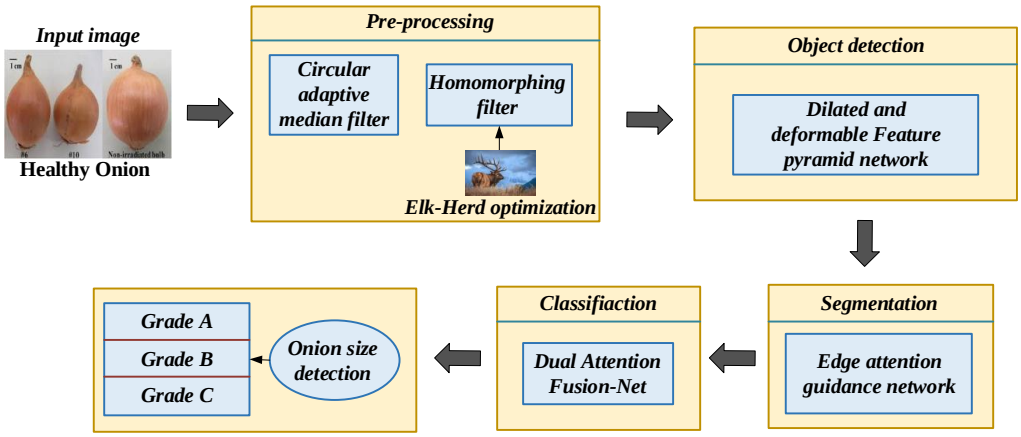


Fig. 1. Process flow of the proposed method.

The workflow of a proposed approach is depicted in Figure 1. The proposed approach consists of healthy onions as input which is gathered from the categorization output such as healthy or unhealthy onions. To enhance the quality of these input images a data augmentation step is applied involving the transformation such as flipping, rotation, zooming, and brightness adjustment .Following that to reduce noise and improve contrast, a homomorphing filter with elk-herd optimization and a circular adaptive median filter is used on the gathered healthy onion images. To recognize the onion instances on images, a dilated and deformable feature pyramid network is used. For segmenting the object from an image the detected portions are segmented using the Edge attention guiding network by providing an edge or line on an image to define the size of the onion. The area of the segmented onion are calculated and based on that, the size of the onion are determined as Grade A, Grade B and Grade C. These classified grades are trained and predicted using the Dual Attention Fusion-Network (DAF-Net).

Data Augmentation

Data augmentation for images is a techniques to artificially expand the size and diversity of a training dataset.It includes the flipping images horizontally or vertically, rotating them at different angles scaling or zooming shift them in different direction ,cropping and altering brightness, contrast.

Pre-processing

Pre-processing is the process of converting an raw data collection into a complete one. Circular adaptive median filter and homomorphing filter with elk-herd optimization based approaches are used for pre-processing the raw data in this proposed approach.

Circular adaptive median filter

A circular adaptive median filter is a type of image processing filter used to remove noise while preserving edges and details in an image. Unlike traditional median filters, which use a fixed-size rectangular window to compute the median value of pixels, a circular adaptive median filter (CAMF) modifies the window's dimensions and shape in accordance with the properties of the surrounding image. The suggested CAMF algorithm works as follows:

Step 1: Noise is applied on the original image y after it has been taken. When considering noise, the basic model is shown in equation (1):

$$x = y + 1 \tag{1}$$

The original image, the noisy image, and the additional noise are represented by the variables x, y, and z, respectively. A working window measuring 3*3. With $x[r, c]$ serving as its center, T is considered [17]. The window's centre pixel, x[r, c], determines whether the area is noisy or not. If the pixel is noise-free, it is processed further using the equation (2) below; otherwise, it is left unchanged.

$$T = \{x[r + i, c + j], (i, j) \in w\} \tag{2}$$

Where w consists of five neighbourhood pixels in a circular pattern centered around $x\,[r, c]$, and where $w \in \{(-1,0), (0, -1{:}\,1), (1,0)\}$.

Step 2: If there are one or more noise-free pixels in the (3*3) circular window configured in equation (3).

$$x[r, c] = MED(T) \tag{3}$$

Step 3: Window T1's size is extended to (5*5) when the (3*3) circular window's pixels are noisy.

$$T1 = \{x[r + i, c + j], (i, j) \in w1\} \tag{4}$$

Where $w1$ is made up of twenty-one neighboring pixels arranged in a circle around $x[r, c]$. $w1 \in$ is within $\{(-2, -1{:}\,1), (-1{:}\,1, -2{:}\,2), (2, -1{:}\,1)\}$.

Step 4: If one or more of the pixels in a (5*5) circular window are noise-free.

$$x[r, c] = MED\{T1\} \tag{5}$$

Step 5: If every pixel in the circular window (5*5) has noise, then $x[r, c] = MED\{T\}$.

Step 6: Until all of the image's pixels have been processed, steps 1 to 5 are repeated.

Homomorphing Filter

Image pre-processing uses the homomorphic filter (HF), an image enhancing technique, to decrease the localization inaccuracy brought on by improper illumination. Images are first transformed from the spatial domain to the frequency domain using HF in order to enhance the contrast and brightness of the image [18]. After the filtering process, the images with lower reflectance are returned to the spatial domain. Equation (6) showed every pixel grayscale in an image can be expressed using the Lambertian-reflectance model as the product of its illumination and reflectance intensities.

$$f(x, y) = i(x, y)r(x, y) \tag{6}$$

Where the grayscale value of the image at point (x, s) is represented by $f(x, y)$. The illumination and reflectance functions are denoted by $i(x, y)\ and\ r(x, y)$, respectively. Two components in Equation (6) need a linear separation in order to apply the HF in the frequency domain. The logarithmic transform applied to both sides of Equation (7) can be used to do this.

$$\ln f(x, y) = \ln i(x, y) + \ln r(x, y) \tag{7}$$

Performing the Fourier transform $\mathcal{F}$ is made possible by the divided equation.

$$\mathcal{F}_f(u, v) = \mathcal{F}_i(u, v) + \mathcal{F}_r(u, v) \tag{8}$$

Equation (8) demonstrates that by applying an appropriate filtering transfer function, uneven lighting can be addressed by adjusting the illumination and reflectance contributions in the frequency domain. As the step after the Fourier transform, let $H(u,v)$ be the filtering transfer function.

$$S(u,v) = H(u,v)\mathcal{F}_f(u,v) = H(u,v)\mathcal{F}_i(u,v) + H(u,v)\mathcal{F}_r(u,v) \tag{9}$$

Where the frequency domain filtered image is represented in equation (9) by $S(u,v)$. In order to enhance the high frequency components while suppressing the low parts, the following formula is used to modify a Gaussian high-pass filtering function for the filtering transfer function $H(u,v)$.

$$\begin{cases} D(u,v) = \sqrt{(u-u_0)^2 + (v-v_0)^2} \\ H(u,v) = (\gamma_H - \gamma_L)\left[1 - e^{-c\left(\frac{D(u,v)}{D_0}\right)^2}\right] + \gamma_L \end{cases} \tag{10}$$

Where the distance between (u,v) and the expression for the centered Fourier transform's origin as $D(u,v)$. The constant c regulates the filtering transfer function's slope's sharpness. The cut off frequency is indicated by D_0. The contributions from high frequencies are increased while those from low frequencies are decreased by γ_H and γ_L. These two parameters can be changed to stretch the contrast of images; the former must be bigger than one, and the latter must be between 0 and 1, making $H(u,v)$ a general high-pass filter. The median value of $D(u,v)$ can be used to determine the cut-off frequency, which is typically a challenging decision based mostly on empirical judgment. By applying the inverse Fourier transform operation to convert (u, v) back to the spatial domain, the enhanced image may be obtained in equation (11).

$$f_{\mathcal{F}^{-1}}(x,y) == i_{\mathcal{F}^{-1}}(x,y) + r_{\mathcal{F}^{-1}}(x,y) \tag{11}$$

The logarithmic transform and linear separation, the image must be restored using an exponential operation to $f_{\mathcal{F}^{-1}}$. Obtain the final centroid localization image by doing this equation (12).

$$g(x,y) = i_h(x,y)r_h(x,y) \tag{12}$$

Where the HF-filtered illumination and reflectance components are represented in equation (12) by the variables $i_h(x,\ y)$ and $r_h\ (x,\ y)$. The HF process is now complete, and the images contrast and brightness will be appropriately adjusted. In this homomorphing filter, the values of $i_h$and r_h optimally selected using Elk-herd optimization.

Elk-herd optimization

The elk, commonly known as the wapiti, is the largest species of deer, second only to moose deer and a member of the deer family. Elks are located in the forested regions of North America and Central East Asia, and they frequently prefer warm weather to cold weather [19]. Elks graze on grasses, plants, bark, and leaves; they are not predators. Elks are therefore seen as being at the base of the food chain hierarchy. All the same, elks are powerful creatures capable of sprinting up to 50 km/h on short distances and jumping, especially in times of danger. Elks also have keen senses of smell and hearing.

The Mathematical Approach for Elk-herd optimization for the proposed approach is given below:

Step 1: Initialization

Initialize the reflectance components as the population of elk-herd in this proposed approach. The mathematical formulation for the initialization procedure is mentioned in Equation (13) and (14).

$$i = i_1, i_2, i_3, \ldots i_n \tag{13}$$

$$r = r_1, r_2, r_3, \ldots r_n \tag{14}$$

Where, i and r indicates the reflectance components of the proposed homomorphing filter.

Step 2: Fitness function

The main goal of this proposed approach is to increase the PSNR rate while increasing the contrast of an image. So, maximization of the PSNR is considered a fitness function here. Equation (15) and (16) provides the mathematical statement used to denote the fitness function.

$$Fitness\ function = Maximize\ (PSNR) \tag{15}$$

$$PSNR = 10\log_{10}(R^2/MSE) \tag{16}$$

To define PSNR, the simplest method is to use the mean squared error (MSE). The maximum pixel value of the image in this case is denoted by MAX_I.

Step 3: Updation

The calve ($x_i^j(t)$) largely based on the traits derived from their mother harem and father bull (x^{h_j}), $\left(x_i^j(t+1)\right)$ genes are passed down via each family. Every family's calve $\left(x_i^j(t+1)\right)$ is formed using the traits primarily taken from their mother harem $\left(x_i^j(t)\right)$ and father bull x^{h_j}. The calf is reproduced as shown in Equation (17).

$$x_i^j(t+1) = x_i^j(t) + \alpha \cdot \left(x_i^k(t) - x_i^j(t)\right) \tag{17}$$

The frequency of the inherited traits in the herd $x^k(t)$ from the randomly selected elk, where $k \in (1,2,\ldots,EHS)$, is determined by the variable α, which is a random value between 0 and 1. It should be noted that a greater value of α increases the probability that random elements would participate in the new calf, which improves diversity. If the index of the calf is the same as that of its mother, then it $x_i(t+1)$ inherits the characteristics of its father bull x^{h_j} and mother harem x^j, as expressed in Eq. (18).

$$x_i^j(t+1) = x_i^j(t) + \beta\left(x_i^{h_j}(t) - x_i^j(t)\right) + \gamma\left(x_i^r(t) - x_i^j(t)\right) \tag{18}$$

In the Elk Herd, where $x_i^j(t+1)$ represents the property i of the calf j at iteration $t+1$. The index of a random bull in the current bull set such that $r \in B$ is represented by r, and the bull of the harem j is represented by hj. In the wild, if a mother harem's bull does not adequately protect her, it may occasionally mate with another bull. The random values γ *and* β which fall between [0, 2], decide which attributes are inherited from previously created calves. In the proposed EHO, the coefficients β and γ might indicate important parameters.

Step 4: Termination

The termination process is ended after getting the best enhanced images.

Object Detection

Deep learning-based object recognition offers a quick and precise way to estimate an object's placement within an image. The aim of the computer vision task object detection is to identify and locate things important in an image. The assignment entails locating things in an image, determining their borders, and grouping them into distinct groups. Along with picture categorization and retrieval, it is an essential component of vision recognition. DDPFN is utilized as an object detection technique in this proposed approach.

Dilated and Deformable Feature Pyramid Network (DDFPN)

In the DDFPN, there are three additional modules [20]. The DDC module is designed to impart knowledge about spatial surrounds to deformable objects. From other features and predictions, Semantic information can be extracted by the CFC and CI modules.

Dilated and deformable convolution (DDC)

When applied to different scale deformable objects, the deformable convolution is inadaptive. Both distinct items and their component pieces can have their features captured by the DDC. This shows that spatial context is something the DDC can offer. Consequently, to improve the Feature Pyramid Network (FPN) multi-scale features, create a multi-scale DDC module. Equation (19) describes the Dilated Convolution (DC) y_{DC} output given an input consisting of integrated features x.

$$\begin{cases} y_{DC}(p_0, b) = \sum_{p_n \in R(r)} w^{DC}(p_n).\, x(p_0 + p_n) \\ B(b) = \{(-b,-b),(-b,0),\ldots.,(0,b),(b,b)\} \end{cases} \tag{19}$$

Where w^{Dc} is the dilated convolution's parameter, p_0 is the pixel position, The 3×3 kernel with atrous rate b is denoted by $B(b)$, and p_n is the dilated convolution kernel's pixel. An increased receptive field of features is achieved by employing a high atrous rate. Equation (20) provides a description of the DDC y_{DDC} output.

$$\begin{cases} y_{DDC}(p_0, b) = \sum_{p_n \in R(r)} w^{DDC}(p_n).\, x(p_0 + p_n + \Delta p_n) \\ B(b) = \{(-b,-b),(-b,0),\ldots,(0,b),(b,b)\} \end{cases} \tag{20}$$

The deformable displacement is represented by Δp_n. This displacement serves as an offset for the standard convolution, enabling the generalization of different aspect ratio, scale and rotation transformations. The dense spatial transformations are handled collaboratively by the DDC module in a thorough end-to-end way.

$$y_{MS-DDC}(p_0, c) = [y_{DDC}(p_0, c, b)]_{b=2,3,5} \tag{21}$$

Where c is the channel index and [.] is the concatenation operation.

Cross Feature Correlation (CFC)

Build a semantic context network, where the correlation between features is the edge and the semantic meaning is the node. The CFC module has three procedures: skip connection, collective summary, and attention extraction. This is the CFC module's output is expressed in equation (22).

$$\begin{cases} y_{CFC}(:, g_i) = y_{MC-DDC}(:, g_i) + y_a \\ y_a = \sum_{g_i} \alpha(g_i) \cdot y_{MS-DDC}(:, g_i) \\ \alpha = w^{CFC} \cdot \left[\sum_{p_i} y_{MS-DDC}(p_i, c)\right]_{c=1,\dots,C} \end{cases} \tag{22}$$

Where the attentive features y_a and the multi-scale DDC features y_{MS-DDC} are connected through the skip connection in the first row. Three categories were created from the multi-scale DDC data, each containing features from a single scale, on the grounds that it was thought that the attentive features could be shared across several scales. The features of each group are enhanced by the attentive features of the skip connection, and the group index is g_i. The second row shows attentive group summation with cross-channel attention α. The third row is where attention extraction is carried out, and w^{CFC} is the FC layer's cross-channel correlation parameter.

Co-occurrence Inference (CI)

The two FC layers make up the forward network that is the CI section. The output undergoes a sigmoid transformation is described in equation (23).

$$z = sigmoid(w_2^{CI}(ReLU(w_1^{CI}\left[\sum_{p_i} f(p_i, c)\right]_{c=1,\dots,C})) \tag{23}$$

Where the two FC layers' parameters are w_1^{CI} andw_2^{CI}. The function for ReLU(.) represents Rectified Linear Units (ReLU), while f (.) represents ResNet features.

Loss of DDFPN

Co-occurrence, location, and categorization losses are all included in the overall loss. The output of CFCs is used to calculate the first two losses, same as in the case of Faster R-CNN. The original Faster R-CNN's head was able to detect objects based on many scale features. Equation (24) computes the phenomenon loss by computing the difference in entropy between the ground truth co-occurrence vector and the predicted phenomenon feature. The ground truth phenomenon vector z^{gt} and the co-occurrence prediction z are supplied to the phenomenon loss.

$$Loss_{co}(z^{gt}, z) = -\sum_{i=1}^{N} z_i^{gt} \log(z_i) + \left(1 - z_i^{gt}\right) \log(1 - z_i) \tag{24}$$

Where N is the number of categories in this case. Add the three losses to train the DDFPN to learn the context-enhanced representations adaptively.

Segmentation

Through the crucial process of extracting meaningful information from visual input, image segmentation enables computers to understand and comprehend images in a way that is comparable to human perception. The suggested approach utilized the Edge Attention Guidance Network technique as a segmentation method for locating and evaluating the edges of an image.

Edge attention guidance network

Edge-attention representations are used by the Edge-attention guidance Network (ET-Net) to guide the segmentation process. Which has the Edge Guidance Module (EGM) and Weighted Aggregation Module (WAM) modules attached at the end and is mostly based on an encoder-decoder network [21]. Figure 2 shows our Edge attention guidance network architecture, which consists of a weighted aggregation module and an edge guiding module in addition to the primary encoder-decoder network. "Conv" stands for the convolutional layer, and "U," "C," and "+" denotes the addition, concatenation, and upsampling layers respectively.

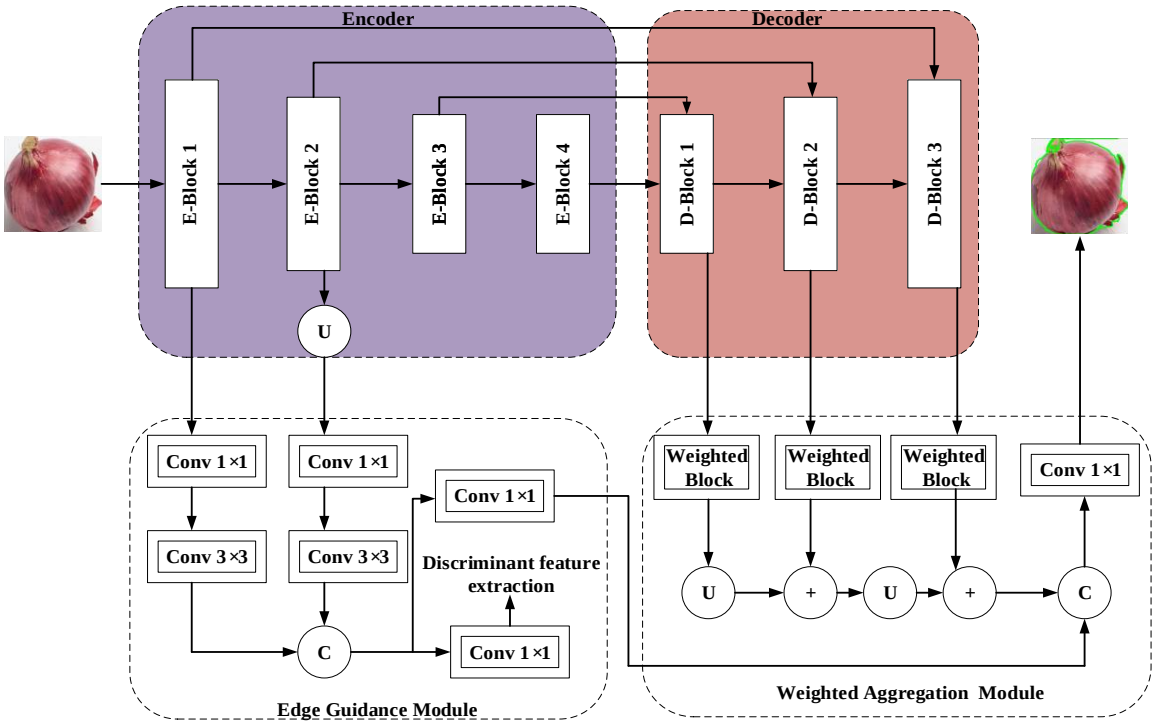


Fig. 2. Architecture of Edge attention guidance network.

Edge Guidance Module

EGM process feeds the outputs of E-Block 2 into the $1 \times 1 - 3 \times 3$ convolutional layers, where they are concatenated after being upsampled to the same resolution as the outputs of E-Block 1. The combined features next go through one of two branches: either an additional 1×1 convolutional layer to predict edge detection results for early supervision, or another 1×1 convolutional layer to act as the decoding path's edge guiding features. EGM employs the Lovasz-Softmax loss, as it outperforms the cross entropy loss in situations involving class imbalance. It can be expressed as follows in equation (25):

$$\mathcal{L} = \frac{1}{C}\sum_{c\in C} \overline{\Delta J_c}\left(m(c)\right), and\ m_i(c) = \begin{cases} 1 - p_i(c)\ if\ c = y_i(c), \\ p_i(c) \quad otherwise, \end{cases} \tag{25}$$

Where C is the class number, and $p_i(c) \in [0,1]$ represents the projected probability of pixel i for class c, and $y_i(c) \in \{-1,1\}$ represents the ground truth label, respectively. The Lovasz extension of the Jaccard index is denoted by $\overline{\Delta J_c}$.

Weighted Aggregation Module (WAM)

Current methods often final predictions are derived by adding outputs at different scales along the channel dimension in order to accommodate differences in object shape and size. Not every function in high-level layers is activated and helpful for object retrieval, though. In order to improve the segmentation performance, edge-attention representations and combine data from many scales and construct the WAM to highlight the important characteristics. The Weighted Blocks receive the outputs from each D-Block, 1 to draw attention to the crucial data. After aggregating the global context information of the inputs using global average pooling,

Through a bottom-up process, WAM combines features from various scales to create feature maps of different sizes comprise a feature hierarchy. In order to retrieve characteristics using edge-guided settings, additionally, the WAM concatenates the EGM's edge-attention representations before using a $1x1$ convolution. The segmentation loss function in our WAM is the Lovasz-Softmax loss, the same like in the EGM for edge detection. Therefore, $\mathcal{L}_{total} = \alpha \cdot \mathcal{L}_{seg} + (1 - \alpha) \cdot \mathcal{L}_{edge}$, is the definition of our ET-Net's total loss function. $\mathcal{L}_{edge}$ and $\mathcal{L}_{seg}$ stand for the losses associated with edge detection in WAM and segmentation in EGM, respectively.

Classification

Classification is the process of organizing objects into groups or classes based on their similarities and affinities. It expresses the shared characteristics that may exist among a wide range of people. DaF-Net is employed as a classification technique in this proposed method.

Dual Attention Fusion-Network:

Dual attention filter is acquired by combining the saliency part proposal network and part attention filter.

i. Saliency part proposal network

An overlaying subnet behind the main network to forecast an object's saliency parts is called a saliency part proposal network. Similar to the popular region proposal network (RPN) used for object detection, SPPN seeks to produce saliency part recommendations [22]. With the help of the location-heatmap, identify the positive samples that are located in the heatmap's activated area. For others disregard the subsequent classification and regression for them and treat them as negative samples. By doing so, SPPN experiences less learning difficulty.

Comparable to regression, classification head, and retina-net are added to each of the three feature map levels in order to forecast the saliency bounding box and convincing score that corresponds to a distinct target size. After NMS, choose the top-N informative regions, then increase SPPN using a self-supervised technique. Similar methods are used by NTS-net, but they save the data exclusively beyond the bounding box and exclude outside data. Regarding that, it might be regarded as a high-level crop in addition to losing the context messages. Claim that maintaining the context message in the interim while restoring the data inside the box makes more sense.

Created a sophisticated way to assess the bounding boxes in the procedure. First, retain each bounding box's pixel from the original input image and each of the N boxes that were selected. In the area that corresponds to the other N-1 boxes, add noise second. Thirdly, apply a strong Gaussian blur to additional locations. The technique obtains N processed images, which it feeds enters the spine to receive N ratings. The implementation of a ranking loss makes the self-supervision mechanism possible. The output will be pushed back into Backbone once more in order to assess the proposal's worth.

ii. Part Attention Filter

The purpose of the attention filter is to attract attention to the aspects of an object that are discriminative. First, an adaptive layer receives the saliency part mask as input. It then outputs a part attention mask that has the same form as the matching feature map. Second, it serves as a filter to draw attention to noteworthy details while ignoring superfluous ones. Draw inspiration from and employ group bilinearity to compute and pairwise interactions overall within the group. Semantic grouping is used to collect feature maps focused on comparable regions. Using these proposed approaches, the onion size is categorized as Grade A (65 mm), Grade B (45 mm) and Grade C (25 mm).

4. Result and Discussion

Detecting and classifying vegetables is a difficult task in everyday production and the difficulty rises when other factors like shape, size, and colour are taken into account. Dual Attention Fusion-Net with Edge Attention Guidance Network based Segmentation are used to onion size detection. Python 3.8.8 is used to assess a desired model on a 64-bit operating system with a 2.50GHz Intel(R), Core(TM) i5-10300H processor, 32.0 GB of RAM (31.8 GB of which is useful), and the following specifications.

Dataset

The real time data was used for training and testing the proposed onions size detection mode. A total of 1000 original images of healthy onions were collected directly from real world agricultural setting ensuring that the dataset authentically reflects practical condition rather than synthetic. These images capture natural variation such as size, differences, natural lighting and organic textures. To enhances the dataset and improves the model ability to generalise data augmentation techniques were applied to the training portion of these real time images. Augmentation methods such as rotation, brightness adjustment, and noise addition and occlusion simulation were used to artificially expand the training set and mimic diverse real world scenarios. The dataset was divided into 80:20 with 800 for training and 200 for testing. In this testing are taken in the process of the confusion model. This ensures the model was both trained and evaluated [23] The unaltered test set included healthy onions in natural settings but lacked onion types, overlapping objects, different growth stages, and environmental factors such as dust, moisture, and changing lighting conditions. Future work aims to expand the dataset with images from farms under diverse conditions, onion types, and environments to improve model robustness and scalability.

Table 1. Simulation Parameter of Elk-Herd optimization.

Parameter	Values
Number of Explore step	3
Number of exploit step	3
Global alpha	0.2
Global beta	0.8
Delta w	2.0

Table 2. Simulation Parameter of Dual Attention Fusion Net.

Parameter	Values
Optimizer	SGD
Activation	ReLU
Epochs	100
Batch Size	64
Loss	Sparse Categorical Crossentrophy

Tables 1 and 2 represent the simulation parameter of elk-herd optimization and dual attention fusion net. Explore step, Exploit step, Global alpha. Global beta and Delta w are the simulation parameters considered for the elk-herd optimization. Dual attention fusion net simulation parameters are optimizer, activation, epochs, batch size and loss.

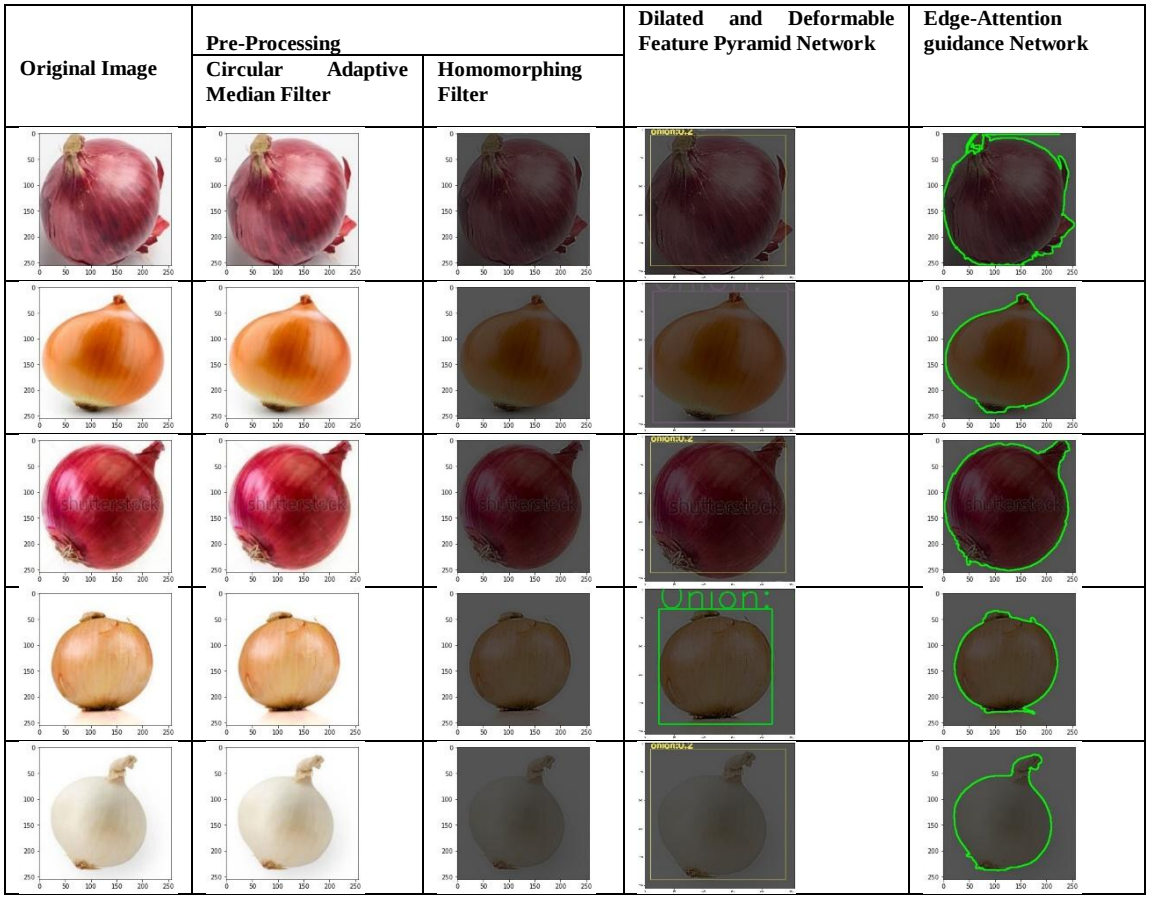


Fig. 3. Sample of pre-processed, object detection and segmentation output.

The figure 3 depicts the pre-processed, object detection and segmentation output of the proposed method. The gathered healthy onion images are used as input images in the proposed approach. Initially the original images are pre-processed in circular adaptive median filter and homomorphing filter in order to eliminate noise and enhance the image quality in order to minimize the localization inaccuracy brought on by irregular elimination respectively. Then, the images are passed into the dilated and deformable feature pyramid network to find the object instances. Finally the resultant image is subjected to edge attention guidance network. In this network, the presence of an edge in an image is detected and the segmented the image in an appropriate way.

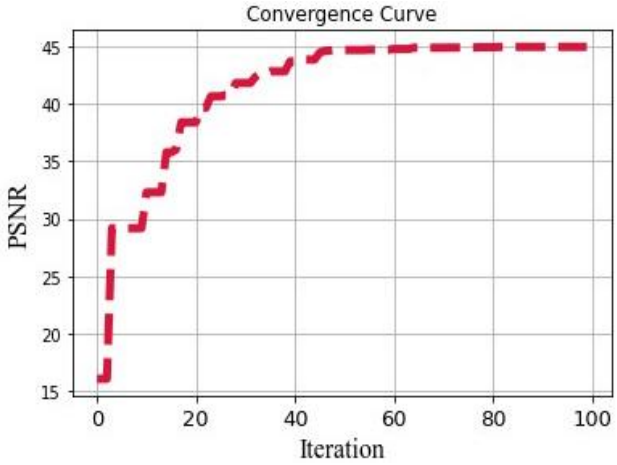


Fig. 4. Convergence Curve.

Figure 4 illustrates the convergence graph for the recommended optimization technique. The PSNR of the proposed optimization is computed using the planned and observed value by the search agents formed for each training set. The curve shows when the iteration is increased at the time PSNR also increased. This figure shows how much better the newly proposed optimization performs than the methods that are currently in use.

Performance analysis of Pre-processing algorithms

The proposed pre-processing approaches for noise removal and contrast enhancement such as CAMF and HF-EHO are compared with other algorithms which are already in use. The comparisons are shown below as a graphical representation.

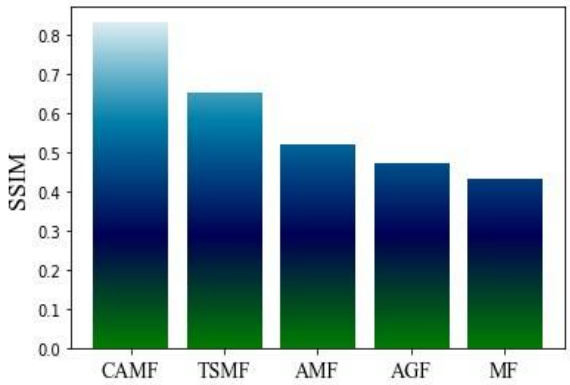


Fig. 5. SSIM Comparison.

Figure 5 depicts the structural similarity index measure (SSIM) comparison for various noise removal algorithm. The SSIM generated by the CAMF, Tri state median filter (TSMF), Adaptive median filter (AMF), Adaptive gaussian filter, and Median filter classifiers are 0.83%, 0.65%, 0.52%, 0.47%, and 0.43%.

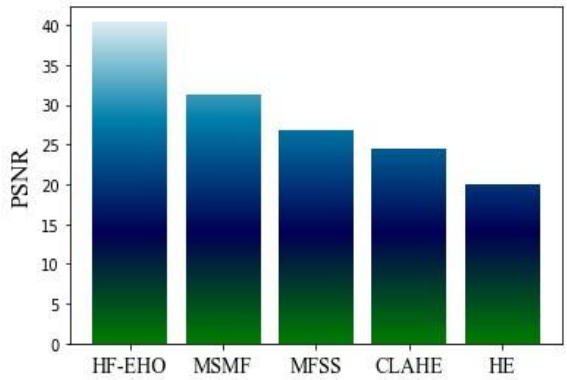
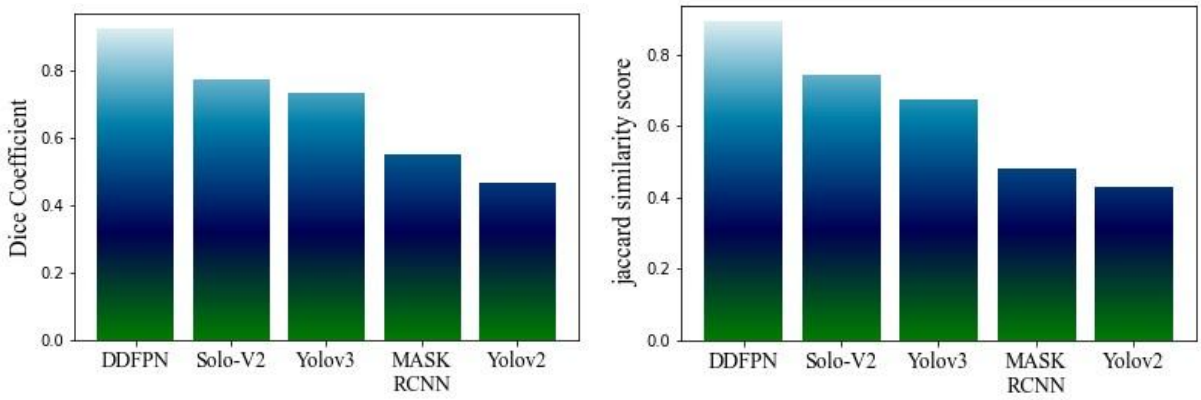


Fig. 6. PSNR Comparison.

Peak signal to noise ratio (PSNR) comparison for different contrast enhancement techniques is represented in Figure 6. The Homomorphing filter-Elk Herd optimization (MF-EHO), the current strategy, has a 40.3% success rate. The other approaches that are currently in use are as follows. Multi-scale morphological filter (MSMF) has a range of 31.2%, Morphological filter single scale (MFSS) has a value of 26.7%, Contrast limited adaptive histogram equalization (CLAHE) has 24.4%, and Histogram equalization (HE) has 19.87%.

Performance evaluation of object detection

The DDFPN utilized for detecting the object in an image is compared with other existing object detection algorithm for evaluating its performance. Dice coefficient and jaccard similarity index are utilized to made the comparison.



(a) (b)

Fig. 7. (a) Dice coefficient, (b) Jaccard similarity score.

Evaluation of various object detection algorithms in terms of Dice coefficient statistic is shown in Figure 7 (a). The proposed DDFPN enhancement algorithm has a high value of 0.92% when compared to other algorithms like Solo-v2, Yolov3, MASK RCNN and Yolov2, which have ratio rates of 0.773%, 0.732%, 0.548% and 0.464% respectively. Figure 6 (b) displays the Jaccard similarity score values for the suggested and current methods. DDFPN is used in the proposed approach, which has a Jaccard similarity score value of 0.891%. However, 0.742%, 0.672%, 0.481% and 0.426% are the current technique's Jaccard similarity score values, such as Solo-v2, Yolov3, MASK RCNN and Yolov2. The Jaccard similarity score value of the suggested strategy is higher than that of the current technique.

Performance analysis of classification

The effectiveness of a proposed DAF-Net is evaluated by comparing it with various existing algorithms. ResNext11, DenseNet201, CNN-BiLSTM and GoogleNet are used for comparison. The graphical representation is shown in the below figures.

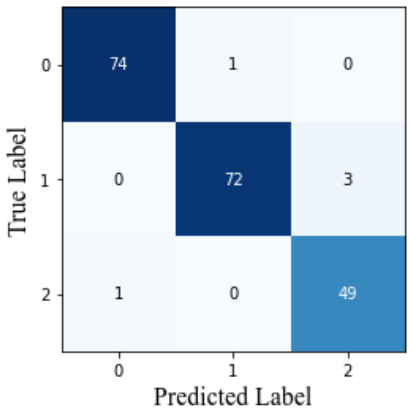


Fig. 8. Confusion Matrix.

Confusion matrix analysis is one technique used to assess a classification system's accuracy. Proposed approach's confusion matrix plot is shown in Figure 8. A confusion matrix is a table that usually shows how well a classification model performs on a set of test data with known real values. Confusion matrices are basic predictive measurements that show recall, specificity, precision, and accuracy. This illustrates how the intended and actual labels in the confusion matrix are misconstrued. The predicted values for 0, 1, and 2 are 74,72and 49 in that order.

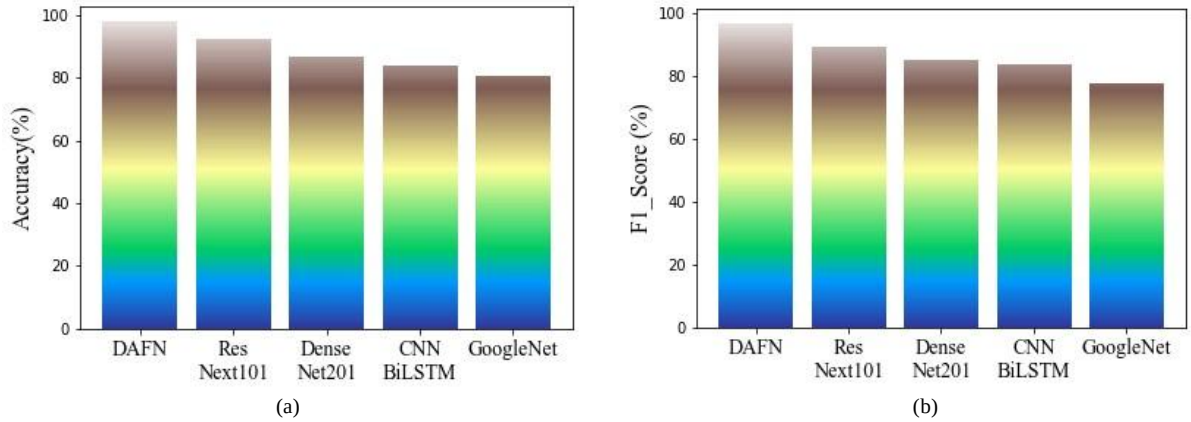


Fig. 9. (a) Accuracy, (b) F1_Score.

The accuracy metrics of the suggested and current methods are contrasted in Figure 9 (a). 97.6% accuracy is achieved with the proposed DAFN. Considering other modern algorithms, it is high. Additional classifiers with an accuracy of 92.4%, 86.8%, 84.1%, and 80.6% are Res Next101, Dense Net201, CNN Bi-LSTM and GoogleNet. The F1-score graphical representation is depicted in Figure 9(b) which has the following existing techniques and the proposed techniques. The value of the DAFN is 96.54%, the ResNext101 has the value of 89.25%, and the Dense Net201 has the value of 84.93%. The CNN Bi-LSTM has a value of 83.55% and the GoogleNet has a range of 77.73%.

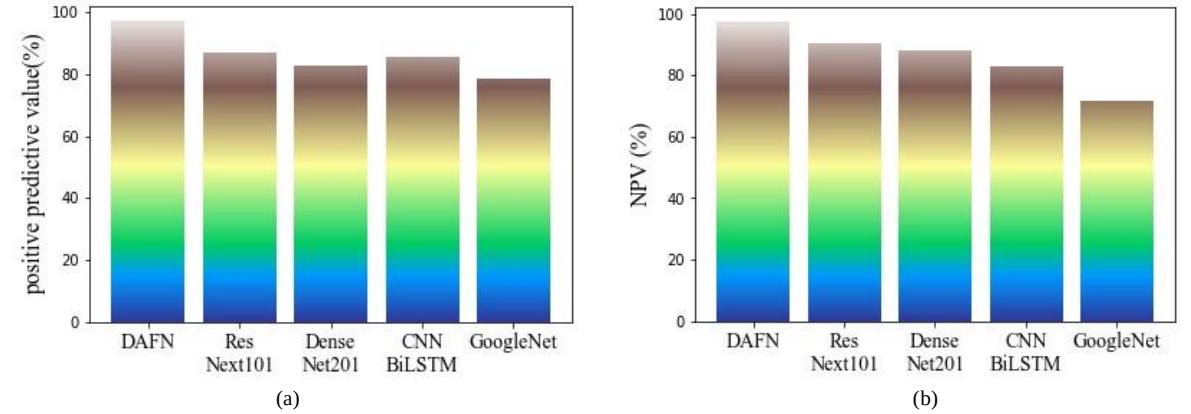


Fig. 10. (a) PPV, (b) NPV.

The Positive Predictive Value (PPV) graphical representation is depicted in Figure 10(a) which has the following existing techniques and the proposed techniques. The value of the DAFN is 97.1%, Res Next101 has the value of 87.2%, and Dense Net201 has the value of 82.69%. The CNN Bi-LSTM has the value of 85.5% and the GoogleNet has the range of 78.7%. The Negative Predictive Value (NPV) graphical representation is shown in Figure 10(b) which has the following proposed and existing techniques named DAFN which has the value of 97.3%, the Res Net101 has the range of 90.6%, the Dense Net201 has the range of 88.3%, the CNN Bi-LSTM has the value of 83% and the value of GoogleNet is 71.8%.

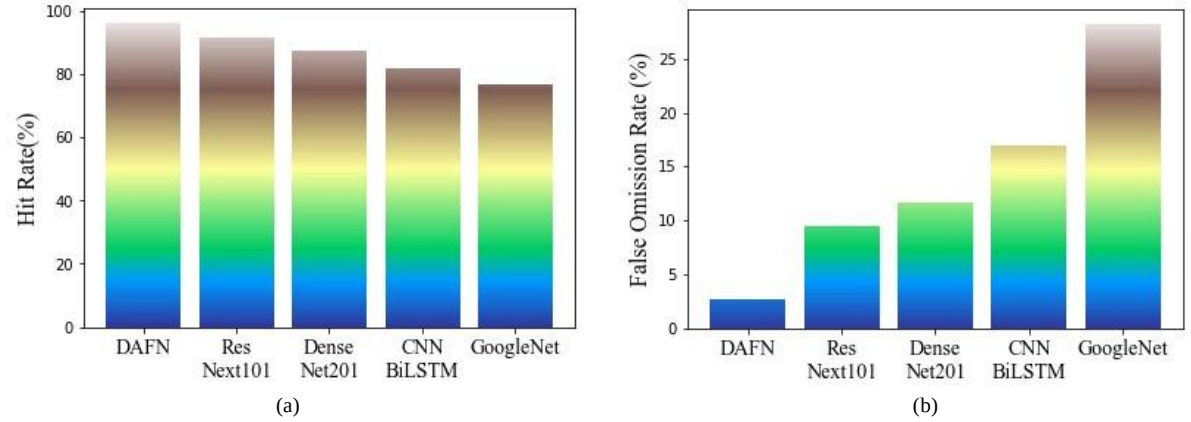


Fig. 11. (a) Hit Rate, (b) FOR.

Figure 11(a) shows the graphical depiction of the Hit rate and includes both the suggested and the current methods. The DAFN value is 96%, the Res Next101 is 91.4%, and the Dense Net201 is 87.3%. The CNN Bi-LSTM range is 81.69%, whereas the GoogleNet value is 76.8%. The False Omission Rate (FOR) graphical representation is depicted in Figure 11(b) which has the following existing techniques and the proposed techniques. The DAFN technique, which is currently in use, has a 2.7% share with the following existing approaches. Res Next101 has a range of 9.39%, Dense Net201 has a value of 11.7%, CNN Bi-LSTM has a value of 17%, and GoogleNet has 28.2%.

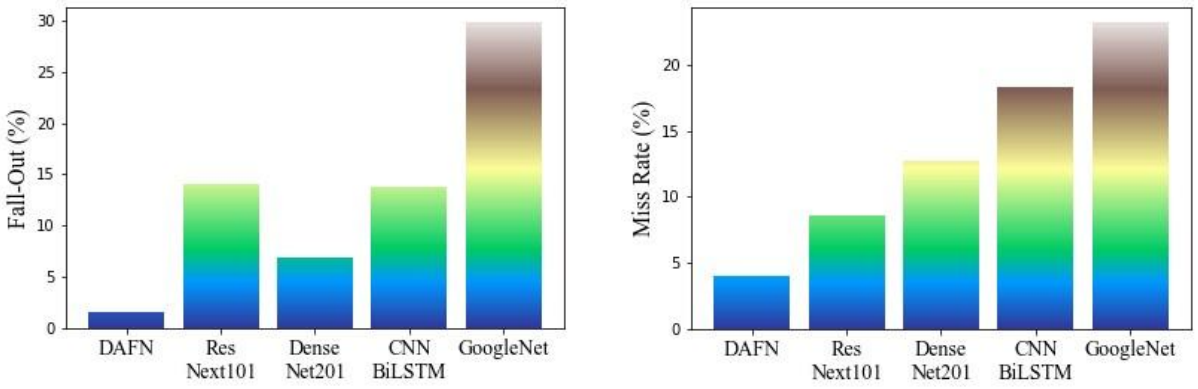
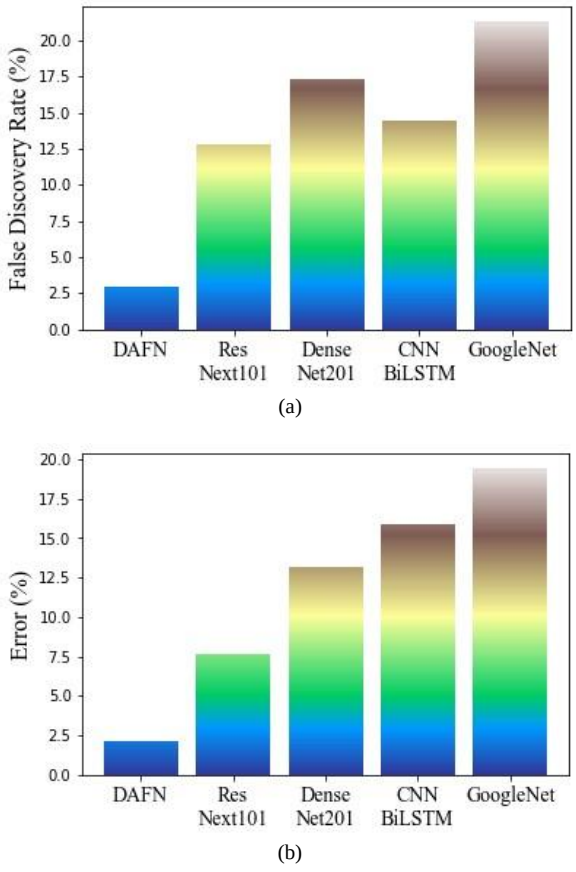
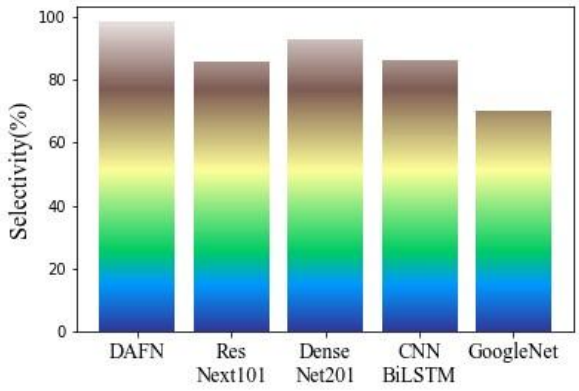


Fig. 12. (a) Fall-Out, (b) Miss Rate.

The Fall-Out graphical representation is depicted in Figure 12(a) which has the following existing techniques and the proposed techniques. The value of the DAFN is 1.5%, the Res Next101 has the value of 14.09%, and Dense Net201 has the value of 6.89%. The CNN Bi-LSTM has the value of 13.7% and the GoogleNet has the range of 29.8%. The Miss Rate graphical representation is depicted in Figure 12(b) which has the following existing techniques and the proposed techniques. The value of the DAFN is 4%, the Res Next101 has the value of 8.59%, and Dense Net201 has the value of 12.7%. The CNN Bi-LSTM has the value of 18.3% and the GoogleNet has the range of 23.2%.

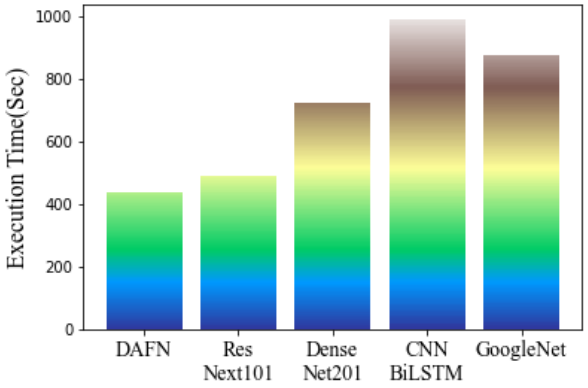




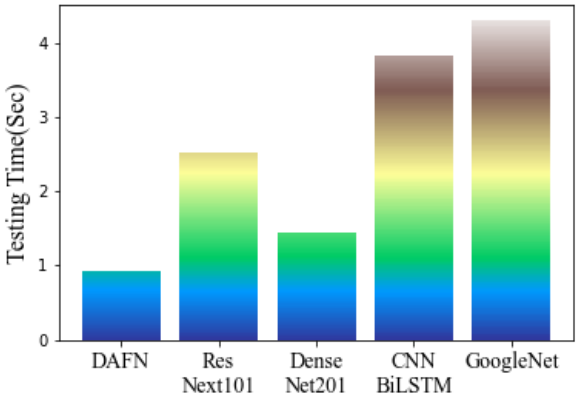
(c)

Fig. 13. (a) FDR, (b) Error, (c) Selectivity.

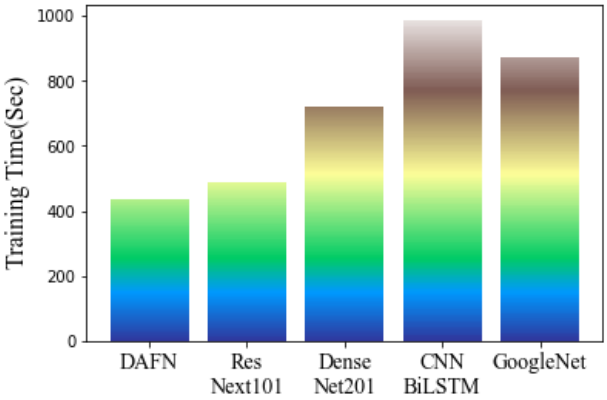
Figure 13(a) compares the False Discovery Rate (FDR) parameter for the proposed and existing approaches. The FDR value of 2.9% for the suggested strategy was found to be greater than the values of 12.79%, 17.3%, 14.5% and 21.29% for the Res Next101, Dense Net201, CNN Bi-LSTM, and GoogleNet approaches, respectively. Figure 13(b) illustrates the error graphical representation using the following suggested and current techniques: the DAFN, with a value of 2.1 %; the Res Next101, with a range of 7.59%; the Dense Net201, with a range of 13.2%; the CNN Bi-LSTM, with a value of 15.9%; and the GoogleNet, with a value of 19.39%. The selectivity graphical representation is shown in Figure 13(c) which has the following proposed and the existing techniques named DAFN which has the value of 98.5%, the Res Next101 has the range of 85.9%, the Dense Net201 has the range of 93.1%, the CNN Bi-LSTM has the value of 86.3% and the value of GoogleNet is 70.19%.



(a)



(b)



(c)

Fig. 14. (a) Execution time (b) Testing time, (c) Training time.

In figure 14(a) shows that the execution time of the algorithm. The time taken in the DAFN is 450 seconds while the existing model such as 490 seconds, 700 seconds, 930 seconds, 850 seconds in the ResNext101, Dense Net201, CNN BiLSTM, Google Net. Respectively. Figure 14(b) represents the testing time. The suggested algorithm is DAFN is 0.9 seconds. The current model is ResNext101 is 2.5, Dense Net201 is 1.5, CNN BiLSTM is 3.9, Google Net is 4.4 seconds. Figure 14 (c) demonstrate the training time. The current model is ResNext101, Dense Net201, CNN BiLSTM, Google Net and the suggested model is DAFN. The values is 470,700,950,850 and the DAFN is 430 seconds Fig. 14.

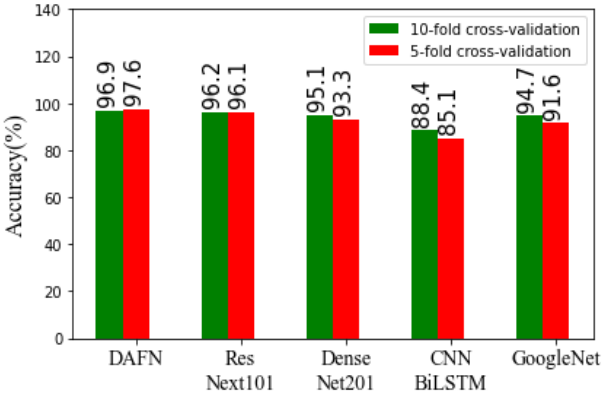


Fig. 15.Cross Validation for the proposed system and existing system.

Figure 15 denotes the cross validation of the model. The green colour shows 10fold cross validation and the red colour shows that the 5-fold cross validation. The 10 and 5 cross validation in the proposed model is 96.9%, 97.6%. Additionally the existing algorithm is 96.2% in the 10 cross validation while the 96.1% in the 5 cross validation in the ResNext101.

Table 3. Cross validation in the system

Fold	Accuracy	selectivity	Miss rate
2	94%	93%	5%
3	96.3%	97.2%	6%
5	97.6%	98.5%	4%

The table presents model performance metrics accuracy, selectivity, and miss rate across different cross-validation folds (2, 3, and 5). As the number of folds increases, the model's performance generally improves. With 2-fold validation, the model achieves 94% accuracy, 93% selectivity, and a 5% miss rate. At 3 folds, both accuracy and selectivity increase to 96.3% and 97.2%, respectively, though the miss rate slightly rises to 6%. The best results are observed with 5-fold cross-validation, where the model reaches 97.6% accuracy, 98.5% selectivity, and a reduced miss rate of 4%. These results indicate that higher fold cross-validation offers a more reliable and accurate evaluation of the model by reducing variance and improving generalization.

Table 4 .Statistical test in the dataset

Features	F-values	P-values
8	1.15	0.3265
16	1.26	0.219
32	1.44	0.053
40	1.31	0.091
48	1.369	0.047
56	1.356	0.041
64	1.317	0.047

The table presents the results of a statistical test conducted on datasets with varying numbers of features, ranging from 8 to 64. For each number of features, the table lists the corresponding F-value and p-value, which together help evaluate the features.

Table 5. Ablation study for the proposed Model

Model	Accuracy	F1-Score	Positive Predictive Value	NPV
CAMF+DAF-Net	94.3%	93%	96%	95%
CAMF+DDFPN+DAE-Net	95.2%	94.4%	93.8%	96.4%
CAMF+ET-Net+ DAE-Net	96.4%	93.9%	94%	93.8%
CAMF+ DDFPN + ET-Net+ DAF-Net	97.6%	96.54%	97.1%	97.3%

The ablation study in table 4 shows that each added module improves the models performances. Starting with CAMF+DAF-Net the model achieves 94.3%accuracy. Adding DDFPN and DAE-Net improves accuracy and F1-score while replacing DDFP with ET-Net raises accuracy further. The best result come from combining all modules CAMF, DDFPN, ET-Net achieving accuracy and 96.54% F1-score.

Table 6. Comparison for the proposed and existing algorithm

Method	Accuracy	F1-Score	MAE
YOLOv8[1]	94%	-	-
R-CNN[2]	-	0.88	5.64
Proposed	97.6%	96.54%	-

The Proposed method outperforms the others with the highest accuracy (97.6%) and F1-Score (96.54%), indicating better precision and recall. YOLOv8 achieves good accuracy (94%) but lacks reported F1-Score and MAE, while R-CNN shows a decent F1-Score (0.88) but a higher error (MAE of 5.64). Overall, the proposed method demonstrates superior accuracy and detection reliability compared to YOLOv8 and R-CNN.

5. Conclusion

Worlds second most widely grown vegetable is onion after tomatoes. 4,562,530 tonnes of onions were produced in 2021 on 215,934 hectares of cropland globally. Although onions are a temperate crop, they may be cultivated in a variety of climates, including tropical, subtropical, and temperate. Onion plants, however, are hardy and can tolerate freezing temperatures while they are young. Onion drought tolerance, however, has been researched in limited detail. Further research is needed to determine the onion's adaption processes. Experiments were carried out using deep learning-based autonomous size measurement of width and height. Onions don't need to be measured by hand; vision technology on mobile devices makes onion classification and packaging simple. This development makes it possible to combine the classification and harvesting processes using AI, which boosts productivity. Their capabilities are not fully utilized in many applications. As onion size detection and harvesting continue to advance, there is potential for significant labour and time savings when using the approaches covered in this paper in real-world applications. In this paper, healthy onions were collected and gathered as a dataset. These datasets are pre-processed for converting the resolution of the images into compatible format. Then, the collected images were pre-processed using the circular adaptive median filter and homomorphing with elk-herd optimization for reduce noise in an image and enhance the contrast level of the noise images. Then, the pre-processed images were object detection based on the location of the objects using the dilated and deformable feature pyramid network. These detected images were segmented using the edge-attention guidance network to determine the presence of a line or edge in an images and outlines. These segmented images were classified using the dual attention fusion-net used to measure the size of the onion and categories the sort's data into labelled classes. The proposed model, the performance metrics such as accuracy, confusion metrics, FDR, Hit Rate and evaluated for both the existing and proposed model. Thus, the proposed onion size detection model is better than the existing model. The performances obtained by this approach were 97.6% accuracy, 2.9% FDR, 96% Hit Rate, 98.5% Selectivity and 97.3% NPV. It is intended to enhance the size estimation model for additional veggies in subsequent research endeavours. In

parallel, it will attempt to enhance the method's long-range performance by utilizing an HD camera and super-resolution reconstruction technologies.

Acknowledgment

The corresponding author claims the major contribution of the paper including formulation, analysis and editing. The co-author provides guidance to verify the analysis result and manuscript editing.

References

[1] M. Mukhiddinov, A. Muminov, and J. Cho, “Improved classification approach for fruits and vegetables freshness based on deep learning,” Sensors, vol. 22, no. 21, pp. 8192, 2022.
[2] P. Mishra, D.N. Rutledge, J.M. Roger, K. Wali, and H.A. Khan, “Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction,” Talanta, vol. 229, pp. 122303, 2021.
[3] S. Karthick, and N. Muthukumaran, “U-Net Based Deep Regression Network Architecture for Single Image Super Resolution of License Plate Image,” Lecture Notes in Networks and Systems, vol. 946, pp. 311-321, 2024.
[4] F.P. Boogaard, K.S. Rongen, and G.W. Kootstra, “Robust node detection and tracking in fruit-vegetable crops using deep learning and multi-view imaging,” Biosystems Engineering, vol. 192, pp. 117-132, 2020.
[5] K.P. Alabi, Z. Zhu, and D.W. Sun, “Transport phenomena and their effect on microstructure of frozen fruits and vegetables,” Trends in Food Science & Technology, vol. 101, pp. 63-72, 2020.
[6] M. Patel, A.Y. Oh, L.A. Dwyer, H. D'Angelo, D.G. Stinchcomb, B. Liu,... and L.C. Nebeling, “Effects of buffer size and shape on the association of neighborhood SES and adult fruit and vegetable consumption,” Frontiers in Public Health, vol. 9, pp. 706151, 2021.
[7] P. Mishra, E. Woltering, B. Brouwer, and E. Hogeveen-van Echtelt, “Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach,” Postharvest Biology and Technology, vol. 171, pp. 111348, 2021.
[8] G.O. Conti, M. Ferrante, M. Banni, C. Favara, I. Nicolosi, A. Cristaldi, and P. Zuccarello, “Micro-and nano-plastics in edible fruit and vegetables. The first diet risks assessment for the general population,” Environmental Research, vol. 187, pp. 109677, 2020.
[9] R. Torres-Sánchez, M.T. Martínez-Zafra, N. Castillejo, A. Guillamón-Frutos, and F. Artés-Hernández, “Real-time monitoring system for shelf life estimation of fruit and vegetables,” Sensors, vol. 20, no. 7, pp. 1860, 2020.
[10] I.S. Minas, F. Blanco-Cipollone, and D. Sterle, “Accurate non-destructive prediction of peach fruit internal quality and physiological maturity with a single scan using near infrared spectroscopy,” Food Chemistry, vol. 335, pp. 127626, 2021.
[11] O.E. Apolo-Apolo, J. Martínez-Guanter, G. Egea, P. Raja, and M. Pérez-Ruiz, “Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV,” European Journal of Agronomy, vol. 115, pp. 126030, 2020.
[12] J. M. Ponce, A. Aquino, B. Millan, and J.M. Andujar, “Automatic counting and individual size and mass estimation of olive-fruits through computer vision techniques,” IEEE Access, vol. 7, pp. 59451-59465, 2019.
[13] J. Gené-Mola, R. Sanz-Cortiella, J.R. Rosell-Polo, A. Escola, and E. Gregorio, “In-field apple size estimation using photogrammetry-derived 3D point clouds: Comparison of 4 different methods considering fruit occlusions,” Computers and electronics in agriculture, vol. 188, pp. 106343, 2021.
[14] H. Jeong, H. Moon, Y. Jeong, H. Kwon, C. Kim, Y. Lee, and S. Kim, “Automated Technology for Strawberry Size Measurement and Weight Prediction Using AI,” IEEE Access, 2024.
[15] M. Ferrer-Ferrer, J. Ruiz-Hidalgo, E. Gregorio, V. Vilaplana, J.R. Morros, and J. Gené-Mola, “Simultaneous fruit detection and size estimation using multitask deep neural networks,” Biosystems Engineering, vol. 233, pp. 63-75, 2023.
[16] R. Sapkota, D. Ahmed, M. Churuvija, and M. Karkee, “Immature Green Apple Detection and Sizing in Commercial Orchards using YOLOv8 and Shape Fitting Techniques,” IEEE Access, vol. 12, pp. 43436-43452, 2024.
[17] P. Sagar, A. Upadhyaya, S.K. Mishra, R.N. Pandey, S.S. Sahu, and G. Panda, “A circular adaptive median filter for salt and pepper noise suppression from MRI images,” 2020.
[18] S. Dong, J. Ma, Z. Su, and C. Li, “Robust circular marker localization under non-uniform illuminations based on homomorphic filtering,” Measurement, vol. 170, pp. 108700, 2021.
[19] M.A. Al-Betar, M.A. Awadallah, M.S. Braik, S. Makhadmeh, and I.A. Doush, “Elk herd optimizer: a novel nature-inspired metaheuristic algorithm,” Artificial Intelligence Review, vol. 57, no. 3, pp. 48, 2024.
[20] K. Wu, Y. Zhang, Z. Xie, D. Guo, and X. An, “DDFPN: Context enhanced network for object detection,” Future Generation Computer Systems, vol. 124, pp. 133-141, 2021.
[21] Z. Zhang, H. Fu, H. Dai, J. Shen, Y. Pang, and L. Shao, “Et-net: A generic edge-attention guidance network for medical image segmentation. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22 (pp. 442-450),” Springer International Publishing, 2019.
[22] Z. Dong, J. Wu, T. Ren, Y. Wang, and M. Ge, “DAF-NET: a saliency based weakly supervised method of dual attention fusion for fine-grained image classification,” arXiv preprint arXiv:2001.02219, 2020.
[23] Dataset 1: India production of ONION (apeda.gov.in), (2021), Accessed by 08-04-2024.

Authors' Profiles

Mythili M. is a PhD scholar at Dayananda Sagar University, Bangalore. She has completed her B.E. from CIET and ME from Anna University. She is currently working in Vemana Institute of Technology and has teaching experience of 16 years. Her research interests are in the field of Deep Learning in Agriculture.

Dr. Vasanthi Kumari P. earned her PhD degree in Information and Communication Engineering from Anna University, Chennai. Her thesis is on Medical Image Compression. She has a total of 23 years of teaching experience. Currently, working as a Professor in the Department of Computer Applications at Dayananda Sagar University. She has a rich teaching experience at various reputed engineering colleges such as Kumaraguru College of Engineering and Coimbatore Institute of Engineering and Technology. She has served as a professor and dean at Ranganathan Engineering College. She has published 15 papers in international journals and 12 in international conferences, and many papers in national conferences. Her areas of interest are image processing, machine learning, and data structures.

How to cite this paper: M. Mythili, P. Vasanthi Kumari, "Dual Attention Fusion-Net with Edge Attention Guidance Network based Segmentation for an Automatic Size Detection of Onions", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.17, No.6, pp. 90-108, 2025. DOI:10.5815/ijigsp.2025.06.06